Accepted Manuscript

Measuring communication difficulty through effortful speech production during conversation

Timothy Beechey , Jörg Buchholz , Gitte Keidser

PII: S0167-6393(17)30236-4
DOI: 10.1016/j.specom.2018.04.007
Reference: SPECOM 2558

To appear in: *Speech Communication*

Received date: 28 June 2017
Revised date: 28 March 2018
Accepted date: 23 April 2018

Please cite this article as: Timothy Beechey , Jörg Buchholz , Gitte Keidser , Measuring communication difficulty through effortful speech production during conversation , *Speech Communication* (2018), doi: 10.1016/j.specom.2018.04.007

Highlights

- Speech modifications are sensitive and reliable markers of environment complexity.
- Lombard effects operate independently at different linguistic levels.
- Talkers increase utterance and turn overlap durations in more challenging environments, given certain task conditions
- A novel tool for eliciting fluent conversational speech is described.

Running head: MEASURING COMMUNICATION DIFFICULTY

Measuring communication difficulty through effortful speech production during conversation

Timothy Beechey

The HEARing CRC; The National Acoustic Laboratories; Macquarie University

Jörg Buchholz

The HEARing CRC; The National Acoustic Laboratories; Macquarie University

Gitte Keidser

The HEARing CRC; The National Acoustic Laboratories; The University of Queensland

Corresponding author: Timothy Beechey

tim.beechey@nal.gov.au

National Acoustic Laboratories

16 University Avenue

Macquarie University, NSW 2109

Australia

Measuring communication difficulty through effortful speech production during conversation

Introduction

Speech produced in noise (Lombard speech) is characterized by increased vocal effort which is manifested in acoustic changes such as increased intensity, mid-frequency emphasis, higher first formant (F1) frequencies and fundamental frequency (F0). Many studies of Lombard speech have argued that these speech modifications have a communicative basis (Cooke & Lu, 2010; Garnier, Henrich, & Dubois, 2010; Hazan & Baker, 2011; Junqua, Fincke, & Field, 1999; Lane & Tranel, 1971). The communicative view of Lombard speech attributes the speech modifications listed above to talkers' intention to increase the intelligibility of their speech for the hearer, relative to neutral speech, in difficult listening conditions. This is consistent with the fact that communication is an inherently interactive behavior which is shaped by dynamic feedback between interlocutors, and accommodation in response to that feedback (Schober & Clark, 1989). Dynamic feedback and accommodation distinguish conversation from passive listening. These two strategies can help to improve communication by providing interlocutors with opportunities to signal comprehension difficulties and therefore to influence the speech production of their communication partner. For example, Branigan, Catchpole, and Pickering (2011) demonstrated that hearers' comprehension when listening to dialogs was better than when listening to monologues and was maximized when the hearer participated in a dialog. When overhearing a dialog, the hearer benefited from the feedback and accommodation that occurred between talkers. When participating in the dialog the hearer was able to elicit accommodation tailored to their own comprehension difficulties. However, studies of Lombard speech have generally not considered holistic communicative contexts and therefore may not adequately reflect many aspects of realistic communicative interactions. Many Lombard speech studies have not considered speech produced during conversations and therefore have not captured the effects of dynamic feedback and accommodation between interlocutors. For example, Lu and Cooke (2009) measured changes in speech when talkers read sentences to a listener and Junqua et al.

(1999) considered speech directed towards an automated voice recognition system. While Cooke and Lu (2010) and Hazan and Baker (2011) measured acoustic phonetic changes in speech during conversational interactions, both studies separated talkers into different booths, or with an acoustically transparent screen. Such separation removes important aspects of natural interaction, including visual cues and a more general sense of co-location. Relatively few studies have considered speech modifications that occur within conversations between co-located talkers. Notable exceptions include Aubanel, Cooke, Villegas, and Lecumberri (2011) where talkers sat across a table without any visual obstruction as well as studies of visual analogues of Lombard speech (C. Davis, Kim, Grauwinkel, & Mixdorff, 2006; Kim, Davis, Vignali, & Hill, 2005).

In addition, the types of maskers generally employed in Lombard speech studies have either been stationary noise or constructed babble noise. Live competing speech has been employed as a masker (Aubanel & Cooke, 2013; Aubanel, Cooke, Foster, Garcia Lecumberri, & Mayo, 2012; Aubanel et al., 2011) which introduces informational masking and allows for the study of temporal strategies whereby talkers attempt to exploit predictable gaps in competing speech to maximize the intelligibility of their own speech. Lombard speech studies have not, to-date, considered the effects of realistic background noise representing real-world locations where conversations are likely to take place.

Finally, the majority of Lombard speech studies have considered low-level acoustic-phonetic parameters such as vocal level, F0, formant frequencies, spectral tilt and vowel duration. Only relatively few studies have considered factors inherent to conversational interaction such as turn-taking and talker overlaps (Aubanel & Cooke, 2013; Aubanel et al., 2012, 2011). As a result, relatively little is known about how conversational dynamics affect Lombard speech. Consideration of speech modifications at higher linguistic levels is crucial for our understanding of the communicative nature of Lombard speech and how talkers may employ different strategies in different circumstances. For example, a talker may vary their vocal level independently of their speaking rate or they may vary their F0 independently of their turn-taking

behavior. As a result, it is logically possible for talkers to modify their speech in complex and even contradictory ways. Analysis of speech modifications at different linguistic levels may therefore provide a richer understanding of communicative effort than consideration of a single level of behavior. To understand communicative effort, it is informative to consider different ways in which talkers may modify their speech in challenging communication settings. Among other strategies, a talker may modify their speech in terms of: (i) the rate of vocal fold vibration which forms the voiced sound source of speech; (ii) the articulation of speech sounds through the shape and compliance of the vocal tract; (iii) overall vocal level; (iv) rate of production; (v) length and complexity of utterances; or (vi) manner of interaction with their communication partner, such as turn-taking behavior. These modifications reflect vocal behavior at different linguistic levels from low-level acoustic-phonetic changes up to prosodic, syntactic and discourse-pragmatic changes. A comprehensive review of talker strategies is provided by Cooke, King, Garnier, and Aubanel (2014).

The aim of this study was to investigate how talkers modify their speech when communicating in realistic acoustic environments of differing complexity at both the acoustic-phonetic level and the interactive level. As a secondary aim, we sought to investigate the reliability of automated acoustic analyses, rather than manual annotation methods, to determine whether the rapid acquisition of speech effort data could plausibly be employed in clinical settings in the future. It was hypothesized that speech modifications at the acoustic-phonetic level, such as vocal level, F0 and formant frequencies, will follow a different pattern of change than modifications at the interactive level, such as turn-taking behavior. Consideration of such a range of speech modifications provides a richer understanding of communicative strategies employed by talkers than consideration of acoustic-phonetic factors alone. It will be argued that automatically extracted Lombard speech measures at multiple linguistic levels may be used to measure changes in communication difficulty and effort during conversation. This approach may be generalized to measure the effects of other factors such as hearing impairment or cognitive impairment on degree of communication difficulty.

Methods

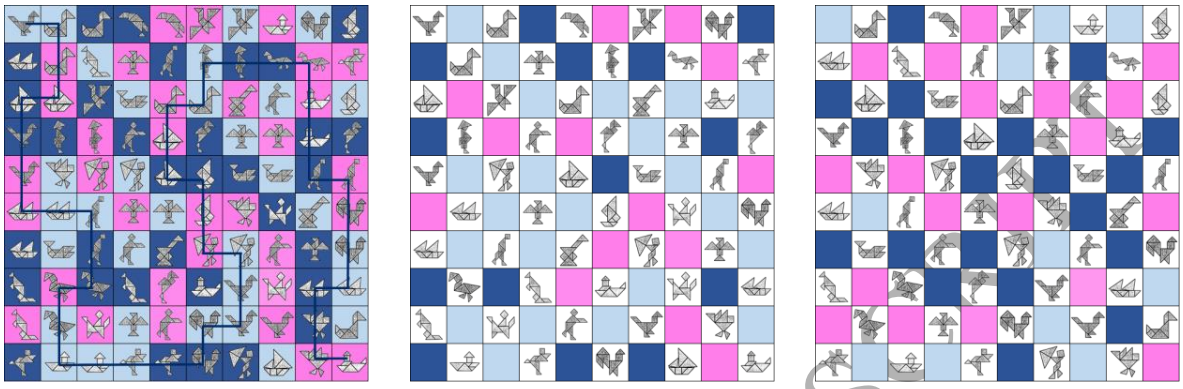
Subjects and materials

Ten male and 10 female native Australian English-speaking adults aged between 18 and 51 years (mean = 28.7 years, standard deviation = 7.97 years) with normal pure-tone hearing thresholds (i.e. < 20 dB HL) between 250 Hz and 8 kHz were tested in pairs. Participants were recruited through advertisements on the Macquarie University campus and through word-of-mouth and received a payment to cover their travel expenses. Participants were naive to the purpose of the study. Treatment of participants was approved by the Australian Hearing Ethics Committee and conformed in all respects to the Australian Government's National Statement on Ethical Conduct in Human Research.

Conversation elicitation task. In order to record fluent, dynamic conversations, a puzzle task was designed to elicit realistically complex utterances and balanced contributions from participants while encouraging engagement. The purpose of the task described here is solely to facilitate fluent, balanced conversations which are as representative of everyday verbal communication as possible. Completion of the task is not a measure of interest as task completion may depend on cognitive resources that are not directly relevant for successful communication. A total of 8 puzzles were constructed on 10 × 10 grids with each square containing a tangram image and one of three colors, which were labeled in the subject instructions as "pink", "dark blue" and "light blue" (Figure 1a). The square colors were chosen to allow for the collection of multiple tokens of the corner vowels from the color names p[ɪ]nk, d[ɐ]rk bl[u]e and light bl[u]e, though analysis of specific vowels is not a focus of the present study.

The object of the puzzles is to find the unique path from the marked start square to the diagonally opposite end square by moving horizontally or vertically between squares containing identical colors or pictures. A single puzzle was created and then 7 additional puzzles were derived with identical structures by rotating and flipping the original puzzle and substituting

different tangram images. The complexity of the puzzles ensured that participants could not detect that the puzzles had a common solution. For each puzzle, two complementary participant views



(a) Combined puzzle (unseen) (b) Participant 1 view (c) Participant 2 view

Figure 1. Example of a complete (unseen) puzzle with solution (left panel) and complementary participant views.

were created by removing half of the information from each square so that every square contained either a color or a picture, but not both. No two adjacent squares contained colors or pictures (Figure 1b and Figure 1c). Participants did not see the complete versions of the puzzles but the entire set of tangram images and colors used in each puzzle was visible in each participant's view. Participants were instructed that they could speak freely and verbally share any information about their puzzle but could not show their puzzle to the other participant. In order to move between any two squares participants needed to share information about the content of the current square and adjacent squares in their puzzle. Participants could not complete any part of the puzzle alone. Tangram images were chosen in order to provide images which were abstract but could be identified unambiguously given sufficiently detailed descriptions. This puzzle task shares similarities with referential communication tasks that have been employed in previous studies but possesses a number of advantages. In comparison to the Sudoku task described by Cooke and Lu (2010) and Aubanel and Cooke (2013), which may be

solved by an individual, the present task necessitates cooperation and communication between participants. In addition, the complexity of the descriptions required to disambiguate tangrams is likely to elicit a more realistic level of grammatical complexity in comparison to the listing of numbers. In addition, the use of a commonly available task such as Sudoku means that participants may have different levels of familiarity with the task which may lead to differences in communication styles when completing the task. Tangram images have been used previously in referential communication tasks. For example, Schober and Clark (1989) and Fox Tree (1999) employed a picture-matching task using tangrams. Unlike picture-matching tasks where one participant holds all the information, the task described here ensures that neither participant holds more information than the other. This ensures that contributions are balanced and that neither participant can opt for a passive role. The Diapix task (Baker & Hazan, 2011; Van Engen et al., 2010) shares similarities with the present task in terms of providing balanced information to participants and encouraging fluent interactions. One difference is that the tangram puzzle task described here is more structured. While the Diapix task requires periods of visual search that may halt conversation, the puzzle task requires participants to find a single path through the puzzle and hence a small set of possible next steps continuously need to be negotiated. It is speculated that this may make it easier for participants in the puzzle task to continue speaking with fewer pauses. The regular intermediate feedback participants received by moving between squares may also facilitate motivation by allowing participants to see the effectiveness of their actions (Nakamura & Csikszentmihalyi, 2001).

The puzzle task was initially tested with two pairs of normal-hearing listeners and was found to meet the requirement of producing a natural and flowing conversation for at least 20 minutes before puzzles were solved. This was later confirmed for all the conversations that were recorded for this study. The task was also found to be engaging to the extent that participants reported forgetting that their conversations were being recorded and expressed a desire to complete the puzzles after the required duration of recordings had been obtained.

Acoustic environments. During each five-minute experimental condition, participants heard one of five acoustic scenes in pseudo-randomized order. Each acoustic environment was played during two experimental blocks for a total of ten blocks. Acoustic environments included a library, an open-plan office, a cafe, traffic on a busy road, and a shopping center food-court. These environments are summarized in Table 1. No silent condition was included as the noise level of the library environment was considered low enough to act as a baseline. All environments were recorded using a purpose-built, 62-channel spherical microphone array (Oreinos & Buchholz, 2016). These environments form a subset of a library of realistic spatial recordings developed at the National Acoustic Laboratories, and were selected because they represented a range of complexities characteristic of common everyday settings. The recordings were transformed into binaural signals by emulating the acoustic path from the microphone array to the ears of a Bruel & Kjaer Head and Torso Simulator using the higher-order Ambisonics sound-field reproduction method. Details are described in Oreinos and Buchholz (2015). The binaural signals were calibrated for playback with Beyerdynamic DT 990 Pro headphones using a GRAS artificial ear (model RA 0045).

Environment	Level (dB A)	Major sound sources
Library	48.5	air-conditioning
Office	56.5	soft speech, laughter, typing
Cafe	76.4	speech, coffee machine, refrigerators
Traffic	79.7	car and truck engine noise
Foodcourt	81.8	speech, music, food production

Table 1

Acoustic environment levels and main sound sources.

Procedure

Subjects were seated facing one-another at a distance of 1.3 meters. Each subject wore a headset microphone (DPA d:fine omni) positioned close to the mouth and open headphones (Beyerdynamic DT 990 Pro). The open nature of the headphones ensured that occlusion of talker's own voices and attenuation of the acoustic path were minimal. The high-frequency attenuation of the acoustic signal was constant across all experimental conditions ensuring that there was extremely limited impact on relative measures. Microphones and headphones remained in place throughout the experiment. To control for the exact position and sensitivity of the headset microphone, each participant was recorded reading from a passage for 30 seconds using the headset microphone as well as a calibrated Bruel & Kjaer (model 4155) measurement microphone positioned 1 meter from the talker at a height of 1 meter. This recording was later used to produce a 1024 taps long, minimum phase, Finite Impulse Response (FIR) filter for each talker to calibrate the level of the speech recorded during the puzzle task. The filter was fitted to the ratio of the octave-band spectra of the speech recorded with the measurement and headset microphone in MATLAB. All speech recordings and signal processing was performed at a sampling frequency of
44.1 kHz.

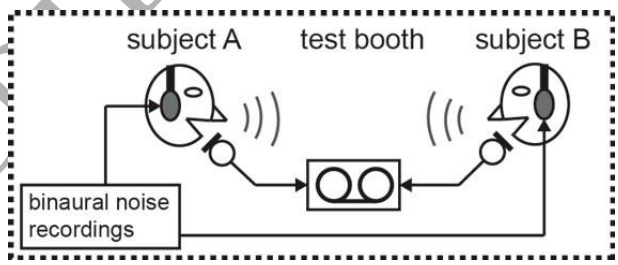


Figure 2. Experimental set-up

Noise was presented through headphones to both participants in 5-minute blocks during which participants jointly solved puzzles, which were attached to a clipboard for convenience. A total of ten experimental blocks were completed for each pair including two presentations of each of

the five acoustic environments. The order of acoustic environments and puzzle versions were pseudo-randomized: the loudest environment was never presented in the first block and all environments were presented once before any environment was repeated. Following each block participants were given a 60 second break. Participants were given the option of ending the experiment at any time if the noise became uncomfortable. None of the participants chose to end the experiment early.

The playback of the noise and the simultaneous recording of the subjects' speech was done on a standard computer running Audiomulch 2.2.4 software and connected to a RME Fireface UC USB sound card and a 4-channel RME Quadmic II microphone pre-amplifier. Because the acoustic environments were presented through headphones they were not picked up in speech recordings. Low level crosstalk was picked up by the microphones, particularly in the louder conditions where talkers produced speech at higher levels. Crosstalk was removed prior to speech analysis.

Subjective listening effort ratings. Immediately following each experimental block, participants rated the perceived level of effort that was required to listen to the speech of their conversational partner. An ordinal rating scale with 13 numerical points from 0 to 12 and with descriptors on the 7 even numbered points was used. The descriptors on even points were 0: "no effort", 2: "very little effort", 4: "little effort", 6: "moderate effort", 8: "considerable effort", 10: "much effort" and 12: "extreme effort" (Johnson, Xu, Cox, & Pendergraft, 2015; Luts et al., 2010).

Speech measure selection

Speech measures were chosen which represent several major components of speech and language production including the source and filter mechanisms of the vocal tract, the length of utterances and the turn-taking behavior of pairs of talkers. Voiced sound source production was measured through talkers' F0. Aperiodic consonant sounds such as stop bursts and frication were

not measured as this analysis would require manual segmentation of the speech signal. Changes in the shape of the vocal tract were measured through the frequencies of the first and second formants (F1 and F2) which primarily reflect widening of the vocal tract through lowering of the tongue and jaw and lengthening of the vocal tract through lip rounding, respectively. Changes in the absorbance of the walls of the vocal tract were measured through the bandwidths of F1 and F2. Decreased compliance of the walls of the vocal tract results in less absorbance of energy, leading to narrower formant bandwidths and greater formant amplitudes. Source and filter mechanisms both contribute to the amplitude of speech which was measured as overall vocal level, and as mid-frequency emphasis, calculated as the difference between the mean level of low frequency critical bands centered at 133, 266 and 531 Hz and mid-frequency critical bands centered at 1060, 2116 and 4222 Hz (Keidser, Dillon, & Byrne, 1998). Interactive behavior was measured through the duration of utterance and talker turn overlaps which reflect conversational organization and turn-taking.

Acoustic analysis

Prior to acoustic analysis, each headset microphone recording was processed to minimize acoustic cross-talk. This was done in MATLAB by first segmenting the two channel (time-aligned) headset microphone recordings separately into 20-ms long time segments using a Hanning window with 50% overlap. The RMS level of each time segment was then compared across the two channels and was kept (i.e., multiplied by a factor of one) only if it was at least 6 dB above the RMS level of the corresponding time segment of the other channel. Otherwise it was removed by applying a multiplication with zero. To avoid the analysis of time segments in which none of the two interlocutors were talking and therefore contained only background noise, time segments with a RMS level below a given threshold were also removed. The threshold was

determined by visually inspecting the recorded speech waveforms before and after the threshold-finding operation was applied, as well as listening to the recordings via headphones. The filters generated during the headset microphone calibration were applied to each single-channel recording to calibrate the individual speech levels. the crosstalk cancellation procedure had the effect of removing brief portions of non-crosstalk speech at the margins of some syllables where the speech envelope had low amplitude. Such deletions primarily affected consonants which have lower amplitude than vowels. There was therefore very little impact on acoustic measurements which were collected from sonorous portions of speech. The introduction of very brief pauses utterance-medially did not affect the segmentation of utterances which treated brief pauses as continuous speech. The trimming of low amplitude margins of envelopes at the beginning and end of utterances had the effect of slightly underestimating overlap durations. Deletion of non-crosstalk speech was greater in louder environments where speech amplitude in the dominant channel was most likely to be great enough to trigger deletion in the non-dominant channel.

Acoustic analyses of speech were carried out using Praat (Boersma & Weenink, 2016). Automated analyses were employed to allow for rapid processing of the speech corpus but were not intended to replace expert manual analyses which are required for any analysis based on semantic content. The use of automated analyses based purely on acoustic information is intended as a first step towards developing a set of measures which may be used in the future to rapidly acquire sensitive and reliable measures of talker effort in a clinical setting. All frequency and level variables (F0, formant frequencies, formant bandwidths and vocal level) were measured within 50 ms windows. This window length was selected to ensure good frequency resolution. Measures for each time window were considered to be valid if finite values for F0, F1, and F2 were found, otherwise all measures for a given window were discarded. Utterance duration was calculated by segmenting each recording into utterances in which intervals containing speech energy with a minimum duration of 75 ms were surrounded by silent gaps of

at least 300 ms duration. This ensured that isolated non-speech sounds such as lip smacks were not parsed as utterances and that brief pauses were not interpreted as utterance boundaries. These specific values were chosen after manually inspecting the duration of non-speech sounds and inter- and intra-utterance pauses in conversation recording across acoustic environments. Talker overlap was measured using a 20 ms sliding window with a 5 ms window increment. Windows with speech energy in both channels were counted as talker overlap. A finite-state parser implemented in the Julia language (Bezanson, Edelman, Karpinski, & Shah, 2017) was used to ensure that brief pauses, including stop closures and medial gaps introduced by crosstalk cancellation, were treated as continuous speech. The parser maintained a state dependent on whether speech energy was present in zero, one or two channels but changed state to reflect absence of speech energy in a given channel only once a pause of at least 180 ms had been encountered. A duration of 180 ms was chosen as being sufficiently long to allow for pauses produced by stop closures and introduced by crosstalk cancellation.

Results of the automated analyses of F0, utterance duration and overlap duration were compared to manual analyses of a subset of the corpus. The middle 60 seconds of conversations in the library (softest), cafe (middle) and foodcourt (loudest) environments for all 10 pairs of talkers were extracted. Within this subset of the corpus, the vowels [5], [I] and [u] were manually segmented from all tokens of the color terms “d[5]rk bl[u]e”, “light bl[u]e” and “p[I]nk”. Visual inspection of pitch tracks in Praat confirmed that all segmented vowel tokens were free of octave jumps. F0 was measured at the temporal midpoint of each segmented vowel. Utterances were manually segmented based on semantic coherence, intonation and pause duration. Non speech utterance such as laughter and coughs, as well as filled pauses were excluded from utterances where they occurred at the beginning or end of an utterance but were included where they occurred medially within an utterance to avoid splitting otherwise coherent utterances. Talker overlaps were manually segmented where utterances in both channels overlapped. Results of manual analyses differed in absolute terms from the automated analysis results but were

proportional in all cases (see Table 2). Proportional agreement between automated and manual analyses indicates that the automated analyses are reliable and not overly affected by artifacts.

	F0 (Hz)		Utterance duration (ms)		Overlap duration (ms)	
	Automated	Manual	Automated	Manual	Automated	Manual
Library	166.98	156.38	2644.14	1689.72	147.17	509.01
Cafe	193.42	190.49	2962.92	2244.96	161.30	617.08
Foodcourt	207.97	203.16	2953.63	2224.03	175.49	774.38

Table 2

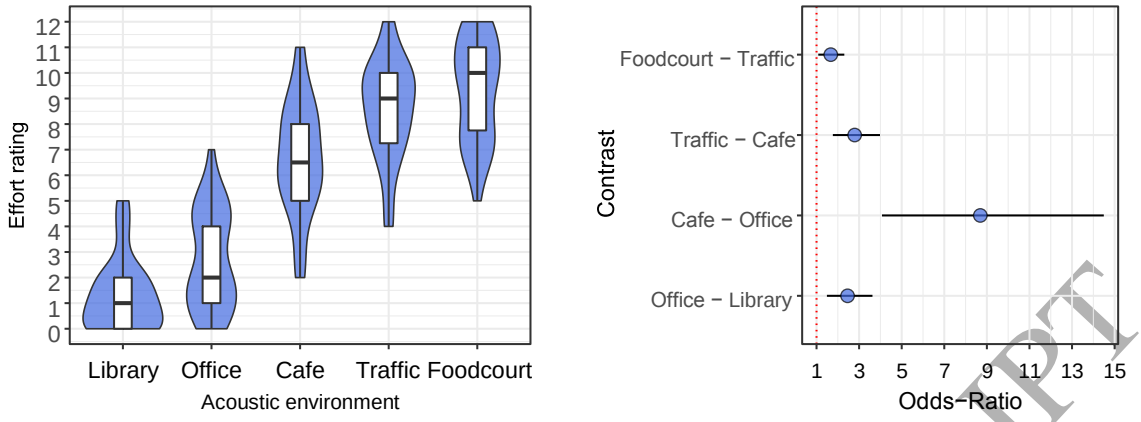
Comparison of automated and manual analyses of F0, utterance duration and overlap duration

All valid measures from the automated analyses were averaged to produce a mean value for each speech measure for each recording resulting in 200 (20 talkers × 5 environments × 2 repetition blocks) observations for each measure.

Results

Subjective listening effort ratings

Effort was rated consistently higher as the SPL of the environment increased, though there is considerable overlap between effort ratings of the traffic and foodcourt environments. The distribution of effort ratings across acoustic environments is shown in Figure 3a. Box plots show medians and interquartile ranges while the width of smoothed kernel densities indicate the volume of ratings in each category.



(a) Summary of subjective ratings of effort on a 13-point scale

(b) Odds ratios of rating in a higher category for pairs of adjacent environments

Figure 3. Subjective listening effort ratings

A Bayesian ordinal logistic regression model was fitted to the effort rating data using the BRMS package (Bürkner, 2017; Bürkner & Vuorre, 2018) within the R statistical software (R Core Team, 2017). An odds ratio (OR) was calculated for each pair of adjacent acoustic environments. The odds of rating difficulty in a higher category increased significantly as the SPL of the environment increased (Figure 3b). Odds ratios greater than 1.0 were found between the library and office OR=2.45 [1.48, 3.62], between the office and cafe OR=8.69 [4.07, 14.49], between traffic and cafe OR=2.79 [1.76, 3.98] and between the foodcourt and traffic environments OR=1.66 [1.08, 2.3]. The odds of higher ratings of difficulty increased the most between the office and cafe environments which likely reflects the relatively large difference in SPL between these two environments.

Speech production measures

Bayesian multi-level models were fitted for each of the speech measures using the R-INLA package (Martins, Simpson, Lindgren, & Rue, 2013; Rue, Martino, & Chopin, 2009) within the R statistical software (R Core Team, 2017). Bayesian methods were selected because: (i) they

allow simple model inference; and (ii) they are valid and robust with small samples sizes as they do not assume asymptotic behavior (Gelman, Carlin, et al., 2014). Gamma likelihood functions were used to model variances which increased with increases in means. Uninformative penalized complexity priors (Simpson, Rue, Martins, Riebler, & Sørbye, 2017) were employed for all models. Models which best fitted the data were selected on the basis of Widely Applicable Information Criteria (WAIC) measures (Gelman, Hwang, & Vehtari, 2014). Final models employed acoustic environment and repetition block group-level effects with random intercepts and random slopes for individual subjects to account for the fact that each talker may have a different baseline and may be affected to a different extent by each environment. F0 was modeled separately for males and females because the pooled male and female data was bimodal due to the difference in average male and female vocal tract size. The distributions of combined male and female data for all other speech measures were unimodal and were modeled jointly. All speech measures show clear patterns of change across acoustic environments with varying degrees of certainty. Mean estimates and 95% CIs for each speech measure are shown in Figure 4. Contrasts across repetition blocks show that the speech production measures are reasonably stable across time (Figure 5). None of the measures were significantly different between repetition blocks, as indicated by 95% credible intervals that do not exclude zero.

Effect sizes

In order to compare the sensitivity of different measures on a common scale, effect sizes were calculated for each of the speech production measures across each of the ten pairwise

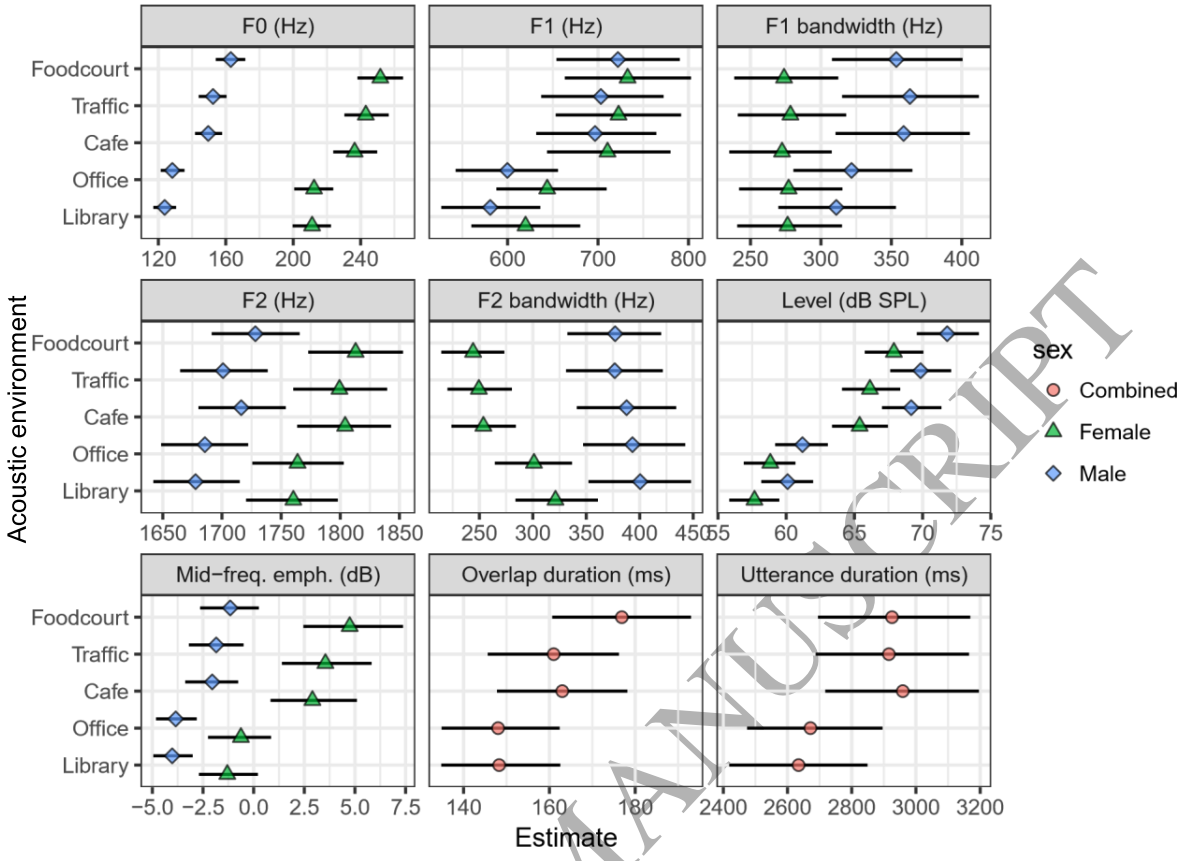


Figure 4. Mean estimates by acoustic environment with 95% credible intervals

acoustic environment contrasts. Effect sizes were calculated using Hedges' g_{av} to account for repeated measures and small sample sizes (Cumming, 2012). For each effect size calculation, the full posterior distributions of credible values of two means were compared. This allows for the straightforward calculation of 95% credible intervals for effect sizes Kruschke (2013). Point estimates for each effect are shown in Figure 6. For reasons of space, credible intervals are not shown.

Combined effect size estimates and 95% CIs are shown in Figure 7 for each environment contrast and in Figure 8 for each speech measure. Effect sizes are classified as small (≥ 0.2),

medium (≥ 0.5) or large ((≥ 0.8) based on conventional thresholds suggested by (Cohen, 1992, 2013).

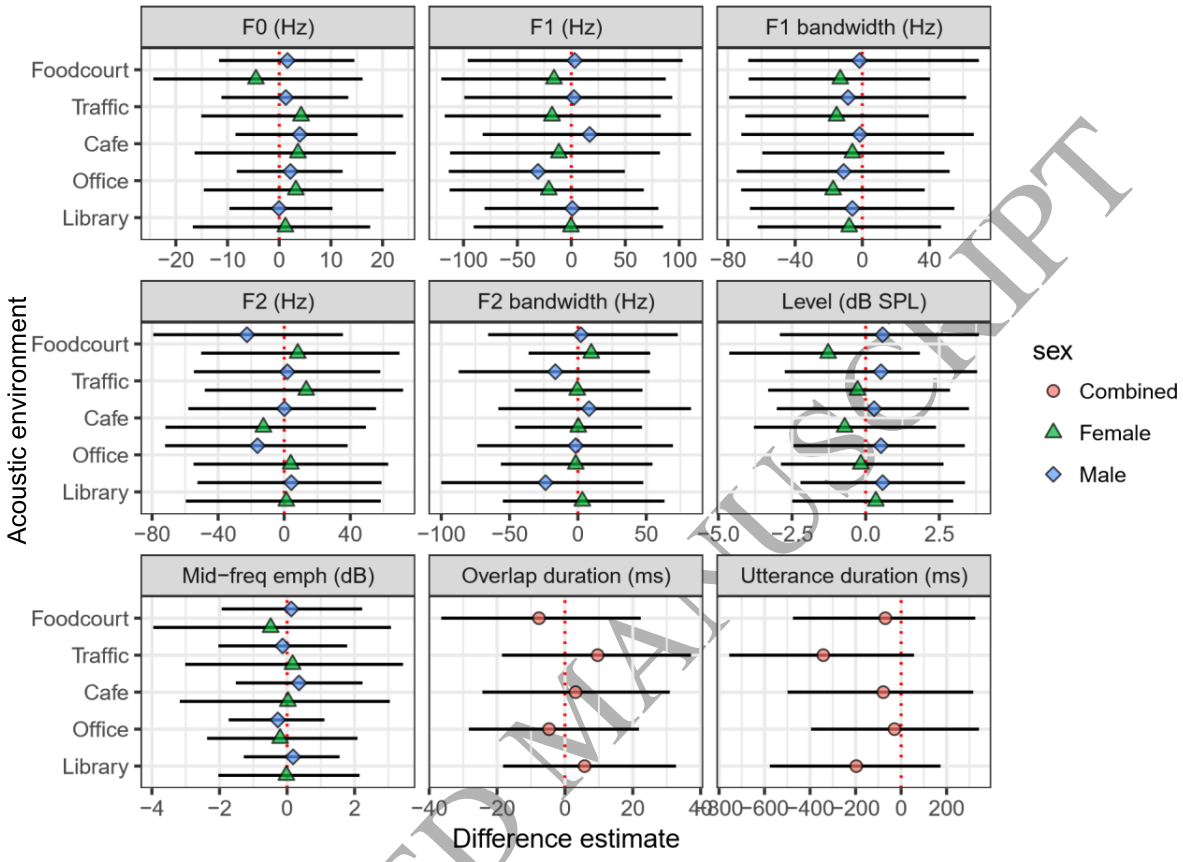


Figure 5. Mean difference of effects with 95% credible intervals for speech measures between the first and second conversation by acoustic environment

Combined effect sizes for the environment contrasts can be broadly grouped into three categories: negligible to small effect sizes for the office-library and traffic-cafe contrasts; reliably small to medium effect sizes for the foodcourt-traffic and foodcourt-cafe contrasts; and reliably large effects for the remaining contrasts between the three loudest (cafe, traffic and foodcourt) and each of the two softest (library and office) environments.

Combined effect sizes for speech measures show the relative sensitivity and reliability of each speech variable. Vocal level and F0 produce large effects across environments for both

male and female talkers. The effect size of mid-frequency emphasis is large for female talkers and medium for male talkers. F1 frequency effects are greater than F2 frequency effects for males and females. For both males and females, the credible intervals for F2 frequency effects extend into the negligible region. The effect size of formant bandwidths is talker sex dependent: for females

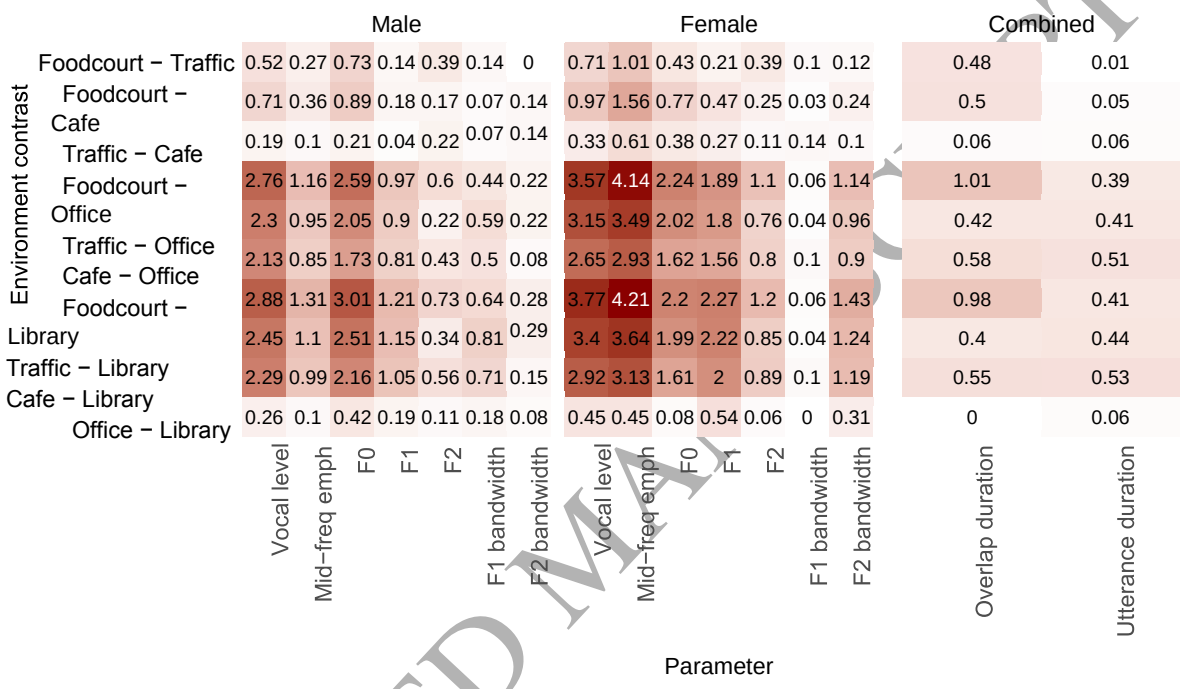


Figure 6. Effect size point estimates. Darker cell colors indicate larger effect sizes.

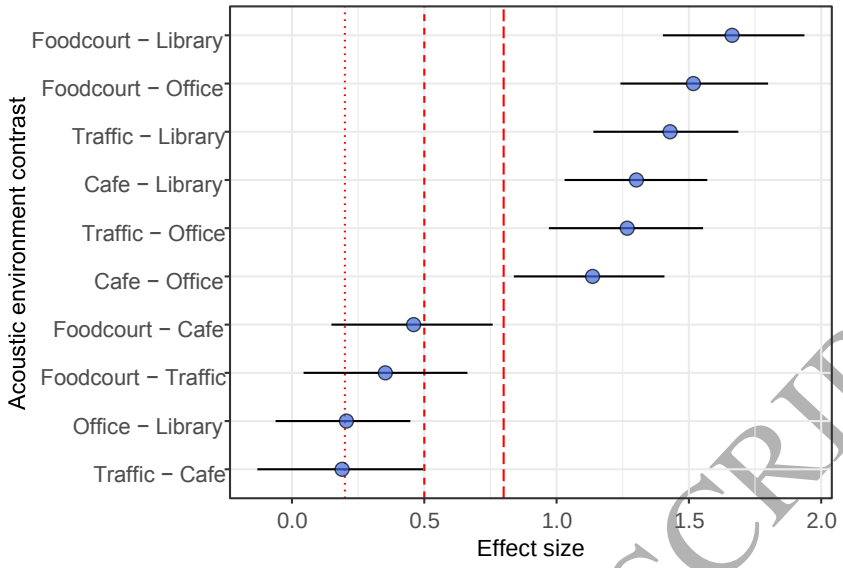


Figure 7. Effect sizes with 95% credible intervals averaged across speech measures for each environment contrast. Broken vertical lines mark the lower thresholds for small, medium and large effect sizes

F2 bandwidth shows a medium effect size while F1 bandwidth is negligible; for male talkers the credible intervals of both F1 and F2 bandwidth extend into the negligible region though the effect size estimate for F1 bandwidth is greater than that for F2 bandwidth. For higher level measures, including utterance duration and overlap duration, credible intervals also extend into the negligible region. In the case of overlap duration, strong effects can nevertheless be observed between the loudest and softest environments, as shown in Figure 6.

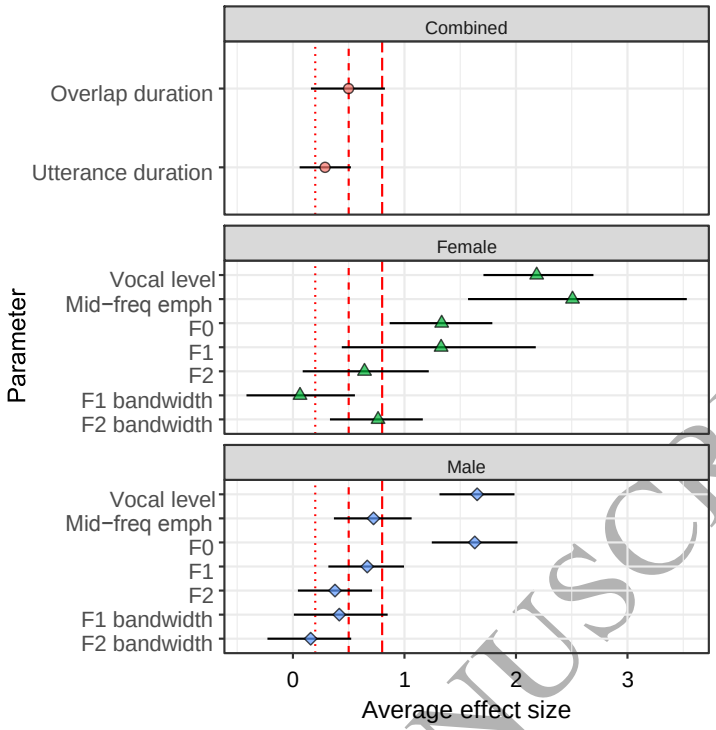


Figure 8. Effect sizes with 95% credible intervals averaged across environment contrasts for each speech measure. Broken vertical lines mark the lower thresholds for small, medium and large effect sizes

Factor analysis

An exploratory factor analysis (EFA) was completed to determine the latent structure of the measured variables. All aspects of the factor analysis were conducted using the Psych package (Revelle, 2017) within the R statistical software (R Core Team, 2017). Parallel analysis (Horn, 1965; Velicer & Jackson, 1990) indicated that 3 factors should be retained. Factors were extracted using the principle axis method which is suitable for data which does not have a multivariate-normal distribution. An obliquely rotated solution was found using the oblimin method. An oblique rotation was chosen to allow for correlations between factors which are expected given that different aspects of vocal production are inter-related. Factor loadings are

shown in Table 3. Absolute values of factor loadings are considered according to conventional thresholds, whereby values greater than 0.6 indicate high loadings, values greater than 0.3 are considered moderately-high loadings and lower values are ignored (Kline, 1994).

	1: Source-Filter	2: Loudness	3: Interaction
Vocal level		**0.475**	0.282
Mid-frequency emphasis	**0.846**	**0.317**	0.156
F1	**0.368**	**0.798**	
F1 bandwidth	**-0.389**	**0.788**	
F2 bandwidth	**0.814**		-0.151
F0	**0.828**	-0.193	0.109
F2	**0.449**	-0.234	**0.331**
Utterance duration	-0.148		**0.576**
Overlap duration			**0.508**
Proportion of variance	0.286	0.187	0.094

Table 3

Factor loadings for an obliquely rotated oblimin solution, with values above 0.3 shown in bold

Factor 1 groups the source and filter components of vocal production. Mid-frequency emphasis, F1 frequency and F1 bandwidth cross-load on factor 1 and factor 2 which groups variables associated with vocal level. Factor 3 groups measures associated with conversational interaction with utterance duration reflecting length of turns and overlap duration representing joint turn-taking behavior. Combined effect sizes for each factor are shown in Figure 9.

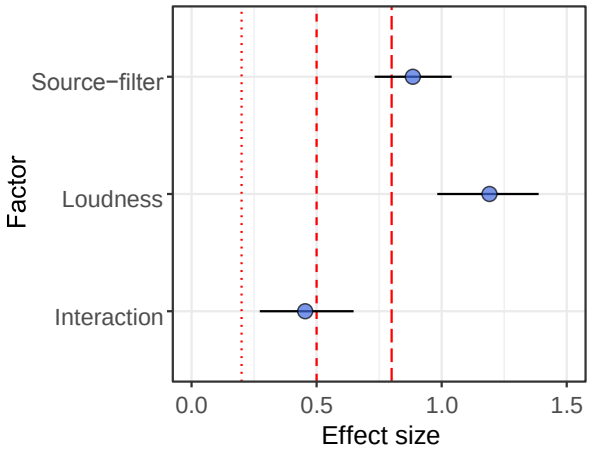


Figure 9. Effect sizes with 95% credible intervals averaged across factors

Note that the magnitude of the effect size of the loudness factor is likely to be more influenced than the other factors by talkers' reduced ability to monitor the level of their own voice inside a noisy environment due to loss of self-feedback in addition to communicative effects. The role of reduced self-feedback may inflate the effect size of vocal level, and of the loudness latent variable, in comparison to other measures.

Discussion

As communication conditions become increasingly challenging, as reflected by subjective ratings of listening effort (section), speech production is characterized by changes of increasing magnitude along multiple dimensions. Talkers increase the frequency of the voicing source of speech (F0) and modify the vocal tract, which is reflected in higher formant frequencies. Both source and filter modifications combine to produce speech at higher levels and with increased energy in the mid frequencies relative to the low frequencies. Spectral changes, reflected in the mid-frequency emphasis measure, were more pronounced in females speech. Changes in formant bandwidths was also dependent on talker sex. Male speech showed increasing F1 bandwidths and no change in F2 bandwidths in louder environments. Female speech, in contrast, showed no

change in F1 bandwidths but decreasing F2 bandwidths in louder environments. In addition to these acoustic-phonetic modifications, talkers produce longer utterances and overlap each other's turns to a greater extent. However, manual review of the recorded conversations revealed no instances of sustained misunderstandings or complete communication breakdown. Participants were able to reliably identify and repair misunderstandings quickly. This is an indication that the speech modifications made by talkers were adequate to maintain successful communication. The characteristics of utterance and overlaps is illustrated in broad transcriptions of 60-second portions of conversation in the library and foodcourt environments in Appendix A.

Factors characterizing speech and communication

The factor analysis reported above indicates that the effortful speech behavior observed in the present study can be grouped into three factors: source-filter, loudness and interaction.

The source-filter factor captures F0, which represents the periodic sound source of voiced speech, and formant frequencies and bandwidths, which represent the resonant peaks formed by the shape of the vocal tract and the degree of damping caused by absorbance by the walls of the vocal tract, respectively. Wider bandwidths reflect lower formant amplitudes. Formant frequencies and bandwidths are expected to load on the same factors because bandwidths tend to widen as formant frequencies increase due to the increased efficiency of absorbance by the walls of the vocal tract at higher frequencies (Stevens, 2000). While this pattern of change was observed for F1 bandwidth in male talkers, F2 bandwidths decreased as F2 frequencies increased in louder environments for female talkers. Narrowing of formant bandwidths may reflect a different strategy employed by female talkers whereby absorbance of the vocal tract walls is decreased to produce higher amplitude F2 peaks. The scale of observed F2 bandwidth changes are likely to be perceptually salient since the bandwidth just-noticeable difference is estimated to be 20-30 Hz with the greatest sensitivity to bandwidth for formants between 1 and 2 kHz (Stevens, 2000). Mid-frequency emphasis loads most heavily on the source-filter factor since it is

the shape of the vocal tract that primarily affects the speech spectrum. Mid-frequency emphasis also cross-loads on the loudness factor as it is a measure of relative levels across the speech spectrum.

The loudness factor combines vocal level with F1 measures. F1 frequency and bandwidth are expected to group with level as increased F1 frequency reflects lower tongue and jaw positions, which increase the size of the resonant cavity and contributes to increased level.

The interaction factor groups overlap duration, the extent to which speakers overlap each other's speech, with overall utterance duration. The pattern of overlap and utterance duration data indicates that talkers may have adopted a "holding-the-floor" strategy in the more challenging communication conditions, whereby talkers speak for longer and compete for speaking turns. That is, speaking obviates the need to listen to, and comprehend, speech. While the task necessitates cooperation rather than competition, talkers could take on more or less dominant roles at different points in time throughout the conversations. In particular, some talkers dominated conversations by asking very specific closed questions rather than more general closed questions and by directing their conversation partner to provide specific information. These findings do not agree with earlier studies which have found that talkers exhibit more careful turn-taking behavior in more challenging conditions (Aubanel et al., 2012, 2011). It is speculated that the difference in tasks may partially account for these observed differences: participants desire to solve puzzles may have motivated them towards goal-oriented behavior and the availability of contextual information in surrounding squares in the puzzle may have made it easier for talkers to take on dominant, guiding roles.

These factors represent three different strategies across two distinct levels of linguistic behaviors. The source-filter and loudness factors represent low-level acoustic-phonetic speech production in terms of the configuration of the vocal tract. Utterance and overlap duration measures reflect turn-taking behavior which is one aspect of the highest level of linguistic behavior: discourse-pragmatics (Levinson & Torreira, 2015). For young, normal hearers, the

source-filter and loudness factors explain most of the variation in the observed data. This supports the hypothesis that lower-level acoustic-phonetic measures are the most sensitive indicators of low to moderate degrees of communication effort.

Effects on intelligibility, comprehension and communication

In conversations between normal hearers, beneficial speech modifications were limited to the acoustic-phonetic level, such as vocal level and mid-frequency emphasis. Note that while some acoustic-phonetic changes have been found not to contribute to improved intelligibility, including F0 (Lu & Cooke, 2009) and vowel duration (Cooke, Mayo, & Villegas, 2014), changes in these parameters are not detrimental to intelligibility. At higher levels, on the other hand, speech changed in ways which could be detrimental to intelligibility and comprehension. For example, longer utterances may increase working memory loads (Gibson, 1998). Finally, increased overlaps may worsen signal-to-noise ratios (SNRs) through increased masking, and either increase cognitive load due to the demands of simultaneously talking and listening or simply reduce attention on listening. It is not clear whether the magnitude of these higher level changes would be detrimental or merely neutral in their effect on communication. It can be concluded, however, that the observed source-filter and loudness modifications were sufficient for normal hearers to achieve successful communication even in the face of potentially detrimental modifications in rate, duration and overlap of speech. In more challenging communication scenarios, such as when one interlocutor is hearing-impaired, beneficial changes may extend to higher linguistic levels. That is, in order to maintain communication, a talker may need to speak more slowly, use shorter sentences, or adopt more careful turn-taking behavior when the addressee is hearing-impaired.

The results presented in this paper reveal a conflict between speech behaviors at different linguistic levels. At the acoustic-phonetic level talkers produce speech with properties that are beneficial or neutral for intelligibility whereas talkers simultaneously modify their speech at higher linguistic levels in a manner that may be detrimental to communication. This supports the hypothesis that Lombard effects at different linguistic levels are independent and that intelligibility-boosting modifications occur first at the acoustic-phonetic level as expected on the view that Lombard effects may operate according to the hierarchical linguistic structure of speech and language (Davis & Johnsrude, 2003).

Calculated effect sizes show that several speech measures are highly sensitive to changes in acoustic environment, at least at a group level, including measures representing each of the linguistic levels considered. Analysis of results across repetition blocks shows that measured speech parameters are also highly repeatable across time. These results support the conclusion that systematic changes in speech production can be used to reliably quantify communicative effort in environments of varying difficulty.

Implications

Speech measures, which are sensitive to changes in the difficulty of the acoustic environment and repeatable across time, may be used to classify the degree of difficulty of communication. The effect sizes reported in section 3.3 show that acoustic-phonetic measures including vocal level, F0 and F1 frequency are most sensitive to changes in the difficulty of the acoustic environment. In addition, higher-level measures including overlap duration and syllable duration were also sensitive to changes in the acoustic scene, but displayed more variability across talkers. While the most highly sensitive measures represent low-level acoustic-phonetic behavior, it is possible that studying the behavior across multiple linguistic levels is more informative. Considering a hypothetical conversation where one participant is hearing-impaired,

it is expected that changes in vocal level observed across different acoustic environments will pattern very similarly to the vocal level results reported here. While in the loudest acoustic scenes more speech effort may be required to accommodate for the hearing impairment, there is a limit to how loud a person can speak over any length of time. When additional changes at the acoustic-phonetic level are no longer possible, the talker would need to employ other strategies, such as producing slower speech or shorter utterances. That is, once a talker has reached their limits using one effort strategy they may need to contribute effort from a different category to maintain effective communication. Consideration of a range of speech measures across linguistic levels may therefore provide a rich source of information regarding communication difficulty.

The speech production data presented in this paper offer a direct measure of communication effort and contain multiple dimensions which may be measured simultaneously and with a high degree of repeatability. While these measures seem analogous to the measurement of listening effort, conversational speech productions represent communication rather than listening. While listening effort assesses the effort required for one-way reception of the auditory signal, it may not account for the demands of active communication. Two-way communication may be both easier and more difficult than listening and is likely to involve the use of cognitive resources over and above those called upon during listening. For this reason, assessing the effort required for listening does not fully reflect the effort required in everyday communication settings. For example, conversational speech is more variable than the type of clear speech sentences often employed in laboratory tests, in terms of rate and linguistic complexity, and is characterized by phonetic reductions and disfluencies which are detrimental to intelligibility and processing. Conversely, conversations provide far greater context and repetition which aids comprehension in comparison to independent sentences (see Mattys, Davis, Bradlow, & Scott, 2012, for a review).

Conclusion

By analyzing an expanded set of speech characteristics including measures not typically considered in studies of Lombard speech, we have shown that Lombard speech is not a uniform phenomenon. Instead, effort is manifested in different, and possibly independent, ways across different linguistic levels. For young, normal-hearing people, acoustic-phonetic aspects of speech production are the most sensitive markers of effortful communication, though changes at higher linguistic levels including prosody, utterance length and turn-taking behavior are also sensitive to changes in the difficulty of the acoustic environment. The robust test-retest reliability of speech production measures reported here indicates that these measures may form the basis of a test of communication effort. Communication effort may be a promising concept for understanding communication ability and disability in realistic settings because it allows us to measure behavior which is directly relevant to communication while accounting for the realistic cognitive demands of conversation.

Acknowledgments

The authors acknowledge the financial support of the HEARing CRC, established under the Cooperative Research Centres (CRC) Programme. The CRC Programme supports industry-led end-user driven research collaborations to address the major challenges facing Australia. The authors thank James Galloway for technical assistance, Katrina Freeston for assistance with research subject recruitment, Robert Cowan for valuable feedback, and the research subjects who participated in this study.

Appendix

Dialog transcription

Within the same 60-second time period the talkers jointly produced 222 syllables in the library environment and 249 syllables in the foodcourt environment. A greater number of talker overlaps occurred in the foodcourt than in the library and the overlaps in the foodcourt were more likely to persist across multiple words, such as lines 4-5, 13-14 and 18-19 in Listing 2. Talkers also produced more disfluecies in the foodcourt than the library, including incomplete words and intra-utterance pauses.

t1: [yup

t2: it's on it it's a different mousy thing though [one of those little yuppie dogs

t1: and it's got sort've weird triangles everywhere and [then a square on top?

t2: [yes

t2: [yes

t1: [yeah I've got that to my left

t2: okay let's go left

t1: yup

t2: I'm light blue n[ow

t1: [okay I've got light blue below me

t2: below okay now I got I got that animal with a flat back and a long neck

t1: it's got like flippers down the bottom or ... oh no ... fla- ah like an "alpaca-y" [sort've deal yup

t2: [yes like an alpaca yeah

t1: okay I've got him below me

t2: okay ... I'm ... dark blue right now

t1: I've got dark blue to the right

t2: mmhmm ... running man

t1: ahh below me

t2: mmkay ... pink

t1: below

t2: running man

t1: ah one more down

t2: one more down hmm dark blue

t1: nuh that's where it ends so've you got running man anywhere else?

t2: I've got running man but then I got pink to the left of running man ... now I've got

t1: ah so below your ... running man

```
27  t2: yeah
28  t1: what color are you there?
29  t2: dark blue
```

Listing 1: Broad transcription of a 60 second portion of a conversation in the library environment.

```
1   t1: oh [so I've got the ship to the right of your ship
2   t2:    [below the ship I got light blue
3   t2: I got light blue
4   t1: yup so I've I've got a ship to the right [so we can go across to the
        right let's go t- there
5   t2:                                          [so we can go right yeah
        so that's where I wa- yeah
6   t1: okay and what color are you there?
7   t2: light blue
8   t1: light blue okay oh sorry so I ... I've got one above
9   t2: okay
10  t1: so we can go up there
11  t2: ok[ay
12  t1:    [alright
13  t1: sorry that [was completely my fault there
14  t2:            [that's alright and then I got I got the thinking man
       above
15  t2: a'right I'm in the thinking man now sorry ... [I got dark blue to
       the right
16  t1:                                               [yep
17  t1: okay I've got the thinking man to the right
18  t2: so we [can go right again
19  t1:       [so we'll go across there yep
20  t2: I've got a ... the sitting man to the right or the thinking man
       below
```

```
21  t1: okay I've got a thinking I'm a thinking man so let's go b[elow
22  t2:                                                          [go
23  t1: ok cool
24  t1: um[m I ... yeah
25  t2:    [okay ... I got da- ah light blue to the right or pink underneath
26  t1: okay umm
27  t2: or I can go light blue to the left
28  t1: I've got a sitting down h- so you're a thinking man aren't you?
```

Listing 2: Broad transcription of a 60 second portion of a conversation in the foodcourt environment. The start of turn overlaps are marked by '`[`' and pauses are marked by '`...`'.

References

Aubanel, V., & Cooke, M. (2013). Strategies adopted by talkers faced with fluctuating and competing-speech maskers. *The Journal of the Acoustical Society of America*, *134*(4), 2884–2894. doi: 10.1121/1.4818757

Aubanel, V., Cooke, M., Foster, E., Garcia Lecumberri, M. L., & Mayo, C. (2012). Effects of the availability of visual information and presence of competing conversations on speech production. In *Thirteenth annual conference of the International Speech Communication Association.*

Aubanel, V., Cooke, M., Villegas, J., & Lecumberri, M. L. G. (2011). Conversing in the presence of a competing conversation: effects on speech production. In *Twelfth annual conference of the International Speech Communication Association.*

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, *43*(3), 761–770. doi: 10.3758/s13428-011-0075-y

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. doi: 10.1137/141000671

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer [computer program], version 6.0.14.*

Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, *26*(10), 1667–1686. doi: 10.1080/01690965.2010.524765

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2018). Ordinal regression models in psychological research: A tutorial. *PsyArXiv Preprints*. doi: 10.17605/OSF.IO/X8SWP

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. doi:

10.1037//0033-2909.112.1.155

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language*, *28*(2), 543–571. doi: 10.1016/j.csl.2013.08.003

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America*, *128*(4), 2059–2069. doi: 10.1121/1.3478775

Cooke, M., Mayo, C., & Villegas, J. (2014). The contribution of durational and spectral changes to the lombard speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *135*(2), 874–883. doi: 10.1121/1.3478775

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Davis, C., Kim, J., Grauwinkel, K., & Mixdorff, H. (2006). Lombard speech: Auditory (A), visual (V) and AV effects. In *Proceedings of speech prosody* (pp. 361–365).

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423–3431.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, *27*(1), 35–53. doi: 10.1080/01638539909545049

Garnier, M., Henrich, N., & Dubois, D. (2010). The influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, *53*(3), 588–608. doi: 10.1044/1092-4388(2009/08-0138)

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman & Hall.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for

Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. doi: 10.1007/s11222-013-9416-2

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. doi: 10.1016/s0010-0277(98)00034-1

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, *130*(4), 2139–2152. doi: 10.1121/1.3623753

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. doi: 10.1007/bf02289447

Johnson, J., Xu, J., Cox, R., & Pendergraft, P. (2015). A comparison of two methods for measuring listening effort as part of an audiologic test battery. *American journal of audiology*, *24*(3), 419–431. doi: 10.1044/2015_aja-14-0058

Junqua, J.-C., Fincke, S., & Field, K. (1999). The Lombard effect: A reflex to better communicate with others in noise. In *Proceedings of the IEEE international conference on acoustics, speech, and signal processing* (Vol. 4, pp. 2083–2086). doi: 10.1109/icassp.1999.758343

Keidser, G., Dillon, H., & Byrne, D. (1998). The change in overall level, spectral shape, and loudness perception of speech produced with different vocal effort. *Australian Journal of Audiology*, *20*(1), 21–31.

Kim, J., Davis, C., Vignali, G., & Hill, H. (2005). A visual concomitant of the lombard reflex. In *Avsp* (pp. 17–22).

Kline, P. (1994). *An easy guide to factor analysis*. Routledge.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. doi: 10.1037/e502412013-055

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research*, *14*(4), 677–709. doi: 10.1044/jshr.1404.677

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731. doi: 10.3389/fpsyg.2015.00731

Lu, Y., & Cooke, M. (2009). The contribution of changes in f0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication*, *51*(12), 1253–1262. doi: 10.1016/j.specom.2009.07.002

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., … others (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *The Journal of the Acoustical Society of America*, *127*(3), 1491–1505. doi: 10.1121/1.3299168

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics and Data Analysis*, *67*, 68–83. doi: 10.1016/j.csda.2013.04.014

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7-8), 953–978. doi: 10.1080/01690965.2012.705006

Nakamura, J., & Csikszentmihalyi, M. (2001). The concept of Flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (p. 89-105). Oxford University Press.

Oreinos, C., & Buchholz, J. M. (2015). Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones. *Journal of the Acoustical Society of America*, *137*(6), 3447–3465. doi: 10.1121/1.4919330

Oreinos, C., & Buchholz, J. M. (2016). Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids. *Journal of the American Academy of Audiology*, *27*(7), 541–556. doi: 10.3766/jaaa.15094

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria. Retrieved from `https://www.R-project.org/`

Revelle, W. (2017). psych: Procedures for psychological, psychometric, and personality research

[Computer software manual]. Evanston, Illinois. Retrieved from `https://CRAN.R-project.org/package=psych` (R package version 1.7.3)

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series b (statistical methodology)*, *71*(2), 319–392. doi: 10.1111/j.1467-9868.2008.00700.x

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. doi: 10.1016/0010-0285(89)90008-x

Simpson, D. P., Rue, H., Martins, T. G., Riebler, A., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. doi: 10.1214/16-STS576

Stevens, K. N. (2000). *Acoustic phonetics*. MIT Press.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and speech*, *53*(4), 510–540. doi: 10.1177/0023830910372495

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate behavioral research*, *25*(1), 1–28.